

# SPECTRAL PARAMETER ENCODING: TOWARDS A FRAMEWORK FOR FUNCTIONAL-AESTHETIC SONIFICATION

*Takahiko Tsuchiya and Jason Freeman*

Georgia Institute of Technology,  
Center for Music Technology,  
840 McMillan St. Atlanta GA 30332, USA  
{[takahiko](mailto:takahiko), [jason.freeman](mailto:jason.freeman)}@gatech.edu

## ABSTRACT

Auditory-display research has had a largely unsolved challenge of balancing functional and aesthetic considerations. While functional designs tend to reduce musical expressivity for the fidelity of data, aesthetic or musical sound organization arguably has a potential for representing multi-dimensional or hierarchical data structure with enhanced perceptibility. Existing musical designs, however, generally employ nonlinear or interpretive mappings that hinder the assessment of functionality. The authors propose a framework for designing expressive and complex sonification using small timescale musical hierarchies, such as the harmony and timbral structures, while maintaining data integrity by ensuring a close-to-the-original recovery of the encoded data utilizing descriptive analysis by a machine listener.

## 1. INTRODUCTION

The long-standing dilemma in data sonification research between functional and aesthetic approaches suggests the difficulty of simultaneously achieving an accurate conveyance of information and a complexity or expressivity of the sound output. As functionalists tend to eliminate unnecessary elements in sonification while aesthetic proponents often interpret data subjectively or employ external information for a "metaphorical" mapping [1], finding a middle ground between them seems challenging. Besides such arbitrariness in mapping decisions, many musical sound organization principles, such as scaling and quantization of the pitch and time, typically require non-linear transformation of data that may result in the loss of information. Despite these challenges, music, as organized sound [2], arguably possesses a potential for multi-dimensional mapping of data optimized for human perception with unique hierarchical sound organizations. We investigate the possibility of this multi-dimensional and higher-order expression for data sonification with enhanced perceptibility while acknowledging the risk of information loss.

### 1.1. Musical Structures

The concepts of musical structures are extensive and differ amongst music theories, compositional styles, and musicology or music information retrieval (MIR) research. In the most ordinary case with western symbolic notation, a musical event,

for example an onset, may have pitch, volume, timing, duration, and timbre as parameters. The pitch may have higher-level structures including harmony, scale, or melodic patterns, while the volume and timing may contribute to the forming of timbre or rhythm. Such hierarchical relationships are not limited to particular styles of music such as Western classical music, but, as analyzed by Roads [2], may apply with varying degree to any aesthetic sound expressions as they all derive from a single time-domain acoustic phenomena. Musical structures may also derive from, besides the level of time resolution, various non-linear data transformations such as scaling (e.g., in range and distribution), quantization, alignment of continuity or discontinuity, dichotomous balances such as repetition and variation, tension and release, and noisiness and tonality. In this paper introducing our initial version of the framework, we limit the focus of musical structures to smaller time-scale hierarchies centered around a common musical event (i.e., note) in relation to the frame-level spectral analysis employed in the assessment of data integrity.

### 1.2. Evaluation Frameworks

A major problem in incorporating expressive or aesthetic techniques in sonification design is the general lack of ways to ensure the data transmission in a quantifiable manner. Many existing evaluation attempts with aesthetic sonification are high-level subjective listening tests that are inevitably influenced by subjective listener preferences and by listener fatigue. In the proposal of Sonification Evaluation eXchange (SonEX) [3], Degra et al. discuss the importance of quantified measurements such as accuracy, error rates, reaction speed, and precision measures. However, even though SonEX provides a framework for standardized and reproducible experiments, it does not present cases for how to actually measure the accuracy and error rates in a complex sonification. Another evaluation scheme, called multi-criteria decision aid, proposes quantitative measurement of the design features, such as clarity, complexity, and amenity, in various sonifications [4]. These evaluation schemes put weight on defining a community-based environment for comparative testing of new sonification designs. However, there seems to be few attempts in objective measurements within a single aesthetic sonification system. While subjective listening tests and statistical analysis may reveal individual mapping patterns with perceptual increase or decrease, it is not practical to test the extensive hierarchical relationships in an expressive sonification against varying data sources. Our framework, therefore, focuses on assessing the minimization of measurable loss of information in the process of musical mapping and transformation. Instead of extending the existing qualitative evaluation methods, we propose the introduction of



This work is licensed under Creative Commons Attribution – Non Commercial 4.0 International License. The full terms of the License are available at <http://creativecommons.org/licenses/by-nc/4.0/>

a machine-listening element, found in speech recognition and MIR [5][6]. Although this study does not directly address the measurement of the perceptive quality of musically-structured sonification, which is still of great importance, using the techniques employed in speech recognition or MIR allow for quantitative measurements that are often modeled around human auditory perception [7] such as tonality and noisiness measurements [6]. MIR also works directly on complex and multi-dimensional sound organizations (i.e., music), which is suitable for the musically-complex (e.g., polyphonic, spectrally rich) sonifications that we hope to achieve. Our experiments reported in this paper employ the most deterministic techniques in the encoding and decoding processes to achieve the retrieval of the data.

### 1.3. Sonification Frameworks

For designing functional-aesthetic sonification with an information-fidelity evaluation, we propose a framework with structural analysis and mapping as well as frequency-domain parameter encoding and decoding processes. We call this framework *spectral parameter encoding* (SPE) for musically expressive sonification.

SPE draws insights from prominent sonification frameworks such as parameter mapping sonification (PMS or PMSon) [8], model-based sonification (MBS) [9][10], and audification [11]. With many overlapping strategies with SPE, PMS provides comprehensive techniques for semi-automating (i.e., aiding decision making) the process of mapping unseen data to various audio-synthesis parameters. It presents many considerations for optimizing the data features, such as the dynamic range, for better perception and clearer auditory presentation with relatively simple mapping to common synthesis parameters. Grond and Berger also discuss the artistic applications (i.e., "musification") of PMS, pointing towards various classic work including *Bondage* by Tanaka [12]. *Bondage* applies direct and systematic mapping techniques to musically present photographic images, using audification and spectral filtering. These techniques are further examined in SPE in the later discussion.

The structural mapping stage in SPE, discussed in detail in the following section, could be considered as a simplified form of PMS, reduced to estimating the dynamic range and time resolution of the input data for an informed selection of the mapping target for synthetic or symbolic parameters in music. Certain mapping of the structural and error components make more sense in the design of spectral parameter encoding, while others may be useful only for human perception.

MBS fundamentally differs from both PMS and SPE with the use of interactive physical models that follow expressive natural acoustic phenomena. Somewhat similar to the assumption in SPE that musical sound structure increases the (multi-dimensional) perceptibility, MBS assumes that a well-defined virtual-acoustic system, which we may experience in a real-life natural environment, enables intuitive comprehension of complex high-dimensional data structures. The data points are mapped to the configuration or the initial state of such physical models, rather than altering the sound-producing mechanisms, therefore allowing the reuse of the same model with different data sources. In contrast, since the SPE takes the dynamic range and time resolution into account for musical organization, it benefits from and requires some level of manual examinations for each new design of musically-structured sonification.

Hermann et al. have explored other sonification techniques that are relevant to the SPE framework. For

example, principal curve sonification (PCS) focuses on identifying the hidden structure across multiple data dimensions [13], while SPE attempts real-time or instantaneous extraction of the structure in a single dimension. Also, as SPE exploits the parameters of magnitude-spectrum distribution such as the mean and variance, Hermann et al. have experimented with utilizing multiple frequency bands in their spectral mapping sonification (SMS) for mapping and analyzing EEG data, enabling multi-dimensional sonification with a rich timbre [14]. While this approach may provide feasible bidirectional relationships between the input data and output sound with little channel interference, it constrains the use of polyphonic timbres to isolated frequency ranges, which SPE tries to address.

Lastly, audification is a rather straight-forward technique that maps a vector of data to the time-domain amplitude of audio samples (after some preprocessing such as scaling and filtering, if necessary). It does not involve any mapping of data structure to hierarchical representation in sound, although it may potentially reveal the structural pattern of the data as a temporal-spectral effect. The resulting audio may arguably provide the highest reversibility to the original data, given that we have access to the digital audification data or be able to capture the acoustic signal with high temporal precision with no acoustic interference. In addition to audification, various techniques exist for digitally encoding or embedding non-musical information into digital audio data, particularly in the field of audio steganography [15]. Our framework, in order to focus on perception, aims for encoding data into an acoustic signal that transmits through the air, and decoded by either a human or machine listener, we exclude discussions about purely-digital data encoding and decoding.

The following sections discuss the main components of the SPE framework with the focus on analysis-driven decision making and spectrally-decodable data mapping. First, we present the overview and the configuration of the design process. We then discuss several approaches to aligning musical and non-musical data structures by means of simple analytics. The following section elaborates the strategies and various musical techniques for designing a "reversible" sonification utilizing parameterized magnitude-spectral distribution.

## 2. THE FRAMEWORK OVERVIEW

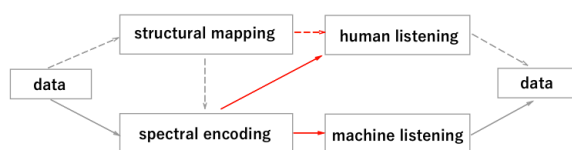


Figure 1: The basic overview of the SPE framework. The solid lines are the essential signal paths for data encoding and decoding while the dotted ones are optional perceptual treatments. Also, the red color indicates an acoustic signal as opposed to a digital signal with the gray color.

The framework for reversible musical data encoding consists of five steps: data input, structural analysis and selection of musical dimension or techniques, spectral encoding, auditory output, and evaluation & data recovery. As shown in Figure 1, some paths are optional and the whole process could consist of only the input, spectral encoding, audio output and evaluation. The structural analysis and mapping may provide additional musical organization to the spectral encoding

process. Both the mapping and encoding stages take techniques of dividing the signal into two distinct components, with somewhat different implications in each stage.

The provided examples in the following discussion are available online<sup>1</sup> for listening and experimenting. The example sonifications utilize a web-browser-based real-time audio environment called Data-to-Music API, developed by the authors [16], which enables various modes of synthesis and sequencing including spectral processing. As the evaluation (or the machine-listening) system is meant to be physically separate from the encoding / sonification system, the online examples do not compute the evaluation results internally. However, the reader may be able to test the examples using common audio descriptor tools, such as the `zsa.descriptor` library for Max/MSP [17].

### 3. STRUCTURAL MAPPING

As stated previously, we are interested in analyzing and repurposing the data structure for musical organization for increased expressivity and perceptibility. By "data structure", we signify not the format of the data organization but the underlying characteristics such as predictable development over time, periodicity, and distribution. SPE takes the simple approach of extracting structures using the additive error model, a commonly employed modeling technique in, for example, data compression, audio and vocal synthesis, and statistical signal processing. The analysis process separates the input signal (data) into rough structural and residual components, such that

$$x(t) = f(t) + \varepsilon, \quad (1)$$

where  $x$  is the input signal,  $f$  is the structural component,  $t$  is the time index, and  $\varepsilon$  is the non-structured component. In a data-compression or audio-synthesis context, the aim of the structural decomposition would usually be the reduction of complex data into more concise parametric representations for further coding or transformation, while retaining the residual part for every data point but in a narrower and more stationary dynamic range than the original form. For the purpose of mapping to musical structures, instead of size reduction, the decomposition of data allows us a flexible mapping of non-musical data to appropriate musical structures without losing the information as a whole.

#### 3.1. Examples of Structural Mapping



Figure 2. An example of decomposition of structure (Left: Input, Center: Estimated structure, Right: Residual signal)

To illustrate the structural mapping process in a very simple sonification problem, suppose we encounter a single-dimensional time series with slowly increasing central values with somewhat stationary noise deviating around them (Figure 2), which may be expressed as

$$x(t) = \frac{1}{1 + e^{-t}} + \varepsilon. \quad (2)$$

Mapping the original signal  $x$  to, for example, the volume of an oscillator (Online Example 1) has a potential of impeding the musical balance or perceptibility as the slow increase of the volume may be hard to hear in the beginning, and it does not take advantage of the dynamic range of human hearing at any given moment. Similarly, if we map  $x$  to the frequency of an oscillator with a fixed amplitude (Online Example 2), although it may represent the data faithfully, it may also produce a sense of "unstable" pitch slowly evolving that does not reside well in more complex sonifications where multiple sound dimensions are presented. The extraction of a larger non-stationary envelope allows repurposing of such non-musical data sources by, for instance, mapping the residue to a full-range of amplitude to hear the detailed fluctuations while the slow-moving central values could be assigned to the frequency to produce more stable and "organized" pitch-sweeping gesture (Online Example 3).

With unordered data, the focus of analysis typically shifts to, for example, clustering, cross correlation, or observing the distribution. In SPE, we utilize the shapes of distribution for musical organization. For example, if a given distribution does not fit to a Gaussian function, we may instead parameterize it into line segments with a fewer number of breakpoints. We can then generate a percussive or metallic timbre with sinusoidal oscillators with the frequency randomly sampled from the parameterized distribution, as described in the next section, while using the residual signal for amplitude modulation (Online Example 4).

Another approach for utilizing the value distribution rather imposes an existing musical structure, such as a musical scale that introduces the minimum amount of distortion, to transform the data non-linearly while retaining the residual values for additional parameter encoding. The Online Example 5 demonstrates the selection of the best musical scale by computing the signal-to-noise ratio after quantizing the data points with the common musical scales in all transpositions. This example may provide improved musical perceptibility, but does not assure a spectrum-based data recovery. For that, the quantization residual signal from may be mapped to, for example, a parameter of the magnitude spectrum, as discussed in section 4.1.2.

#### 3.2. Estimation of Data Structure

To roughly estimate a non-stationary envelope structure in unknown data, we may naively apply a high-order moving-average filter (mocking iterative linear prediction analysis) and then examine the uniformity of the residual noise by calculating the normalized entropy (or variance) of the distribution (Online Example 6). The resulting residual signal is not necessarily uncorrelated (i.e., may contain periodic patterns), but the increase of stationary quality enables more optimal mapping to certain musical parameters. In addition to finding a larger structure, a spectral filter may also capture high-frequency repetitions (e.g., by isolating the smaller coefficients in discrete cosine transform) (Online Example 7). This approach is similar to separating the salient resonances and the noise floor in spectral modeling synthesis [18], in which the noise floor can be replaced with a parameterized spectral noise generator. In addition to these contour

<sup>1</sup> <https://GTCMT.github.io/DataToMusicAPI/icad2017> (Online Examples, Accessed May 15, 2017)

extractions, predictive analysis or the first-order differential of the signal may capture local sequential dependencies in the data.

Although more complex and multi-dimensional statistical analysis using iterative computation may be beneficial, our motivation of using a relatively naive estimation is to allow real-time design (e.g., live coding [19]) of sonification with unseen data, including short-time analysis and mapping of streaming data. The simplicity of the analytical process also helps in retaining intuitive relationships between the input signal and the output parameters, which take relatively close-to-original data structure, while dynamic transformations such as dimensionality reduction may not be suitable for the purpose of the data recovery from the resulting audio.

Therefore, the use of an additive model is beneficial in that it provides a coarse structural component for human perception (and potentially for MIR classification tasks or a model-based estimation of a cleaner signal). On the other hand, the residual component is suited for preserving fine details that may be deterministically recovered by spectral feature extractors [6][17]. Combining the structural and residual parts, after proper rescaling, produces a close-to-original estimate of the input data.

#### 4. SPECTRAL PARAMETER ENCODING

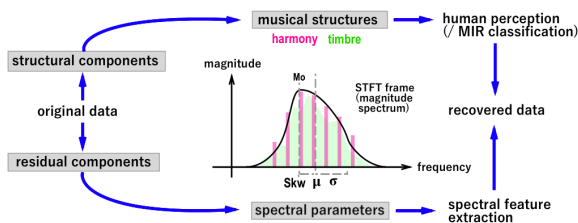


Figure 3. A signal flow incorporating structural analysis (Section 3) into spectral parameter encoding.

The framework aims for not only aligning the musical dynamics for multi-dimensional comprehension, but also preserving as much information as possible in the acoustic signal for computational data retrieval. As shown in Figure 1, this spectral encoding process may be applied directly and entirely to the raw incoming data, or may be combined with a structural analysis to utilize extra information in the mapping and generation processes. A sensible approach may be to route the structural element of the data to the selection of musical expressions to generate, such as harmony and timbre, while using the subtracted residual part for spectral parameters that define the acoustic contour (see Figure 3). However, both signal paths may also be used for spectral encoding, while the musical content is kept as ornamental or non-essential for computational data recovery.

Similar to the additive model used in the structural mapping, the spectral parameter encoding combines two parts, statistical magnitude-frequency parameters and a matching distribution, to generate an acoustic result, such that:

$$y = IFFT[g(\theta_t, X)]. \quad (3)$$

The distribution parameter variables,  $\theta$ , deterministically holds the input data (that are linearly scaled if necessary), which are later measured by spectral descriptors to estimate the original (or post-scaled) values. The encoded parameters may be the first-order descriptive statistics such as spectral

centroid (weighted mean), spectral spread (variance), skewness (median), and spectral crest factor (tonality measure). The actual contents of the spectral distribution do not matter as long as they satisfy the statistical analysis in the short-time Fourier transform (STFT) signal [20]. This allows us to employ various approaches for creating musical expressions, including timbral, harmonic (with musical scales), and polyphonic mixed-timbre voices, that are expanded from the target parameter values.

#### 4.1. Spectral Encoding Techniques

Here, we discuss several techniques of musical sound composition that conform to the acoustic constraints for a sufficient level of data retrieval. The encoding part may utilize either time-domain or frequency-domain synthesis via discrete-time Fourier transform (FT), while the analysis takes part in the FT of the output audio signal.

##### 4.1.1. Timbral Structures

First, we present several of the timbre-based expressions. After specifying the distribution function parameters such as the mean and variance from the input data, one may use spectral filtering to create percussive or sweeping ambience with a changing noise-color (Online Example 8), similar to Tanaka's Bondage discussed previously. This encoding technique requires a frequency-domain element-wise multiplication of the given envelope to the STFT of a white noise, then inverting to the time-domain signal. The resulting audio enables a relatively robust data retrieval with spectral feature description (e.g., centroid and spread) or envelope estimation. Aside from amplitude changes over time such as attack and decay shapes, the spectral content (white noise) may not provide additional structure for perceptual sonification compared to other techniques discussed below. However, the distribution function may take an elaborate shape with linearly-interpolated break points, which would be analyzed with peak estimation or band-limited magnitude analysis similar to SMS. While this approach is efficient for real-time synthesis, it requires the matching of the input vector length and a half of the FT frame by, for example, linear interpolation or zero-padding the extreme frequency ranges.

Though the spectral filtering approach may be good for creating expressions of generic noise percussion, its timbral dynamics can be fairly limited. Instead of using a time-domain noise generator, the oscillator-bank-based approach allows us to take the same spectral distribution function but create more focused (pitched) timbre with non-harmonic partials suited for metallic percussion or ambient pad sounds. This may be realized with a granular (i.e., random-phase, Online Example 9) or an additive (i.e., synchronized-phase, Online Example 10) synthesis using, for example, the random-sampling technique using the inverse transform [21] of the cumulative spectral distribution.

The additive synthesis in the oscillator-bank approach is similarly robust for the recovery of data as spectral filtering. However, the granular synthesis tends to introduce phasing interferences among partials, making the estimation of the statistical parameters less reliable. Since the random time-domain source signal in these timbral-composition approaches cannot be easily estimated, they may be suited for representing the residual noise envelope from the structural analysis step. The structured component, such as a slow-moving contour, may be utilized as gain envelope after normalizing the magnitude spectrum.



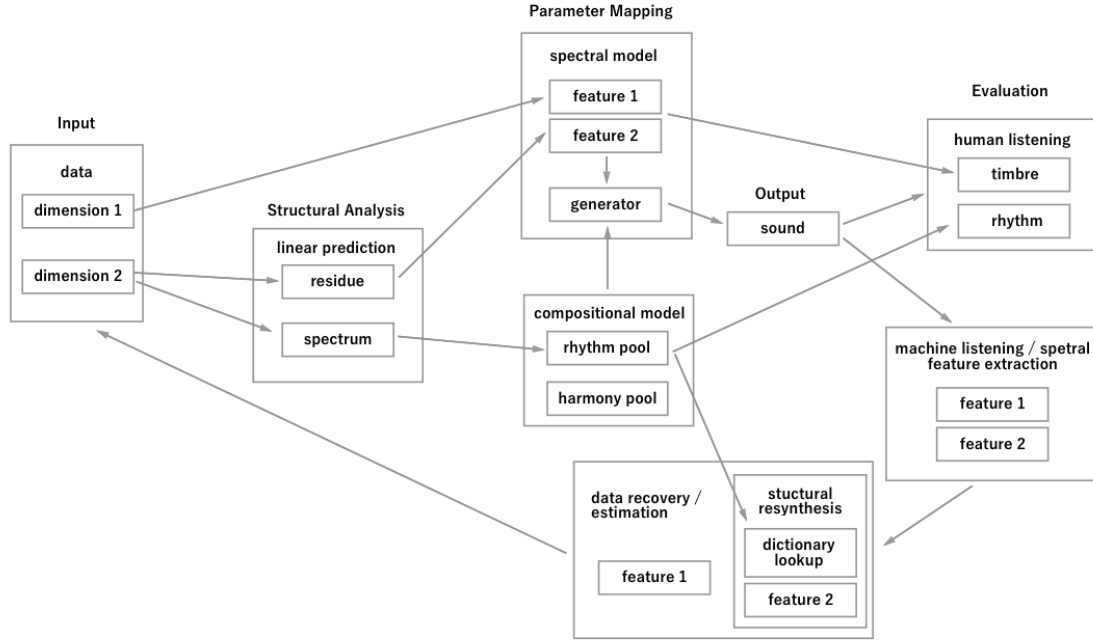


Figure 4. An example of multi-dimensional SPE system

#### 4.1.2. Harmonic Structures

In a more symbolic-level musical organization, one can take spectral parameters, in which we encode data, and generate a single note with natural harmonics or multiple notes forming an arbitrary harmony. For instance, given a spectral centroid (weighted mean), it is trivial to expand it to a single note with  $N$  harmonic partials with a fixed unit amplitude by

$$\begin{aligned} n &= 1, 2, \dots, N; N \in \mathbb{Z}, \\ v_n &= \frac{nN\mu}{N!}, \end{aligned} \quad (4)$$

where  $v$  is the vector of frequencies for sinusoidal additive synthesis,  $\mu$  is the spectral centroid in Hz, and  $N$  is the number of the harmonics (Online Example 11). We can also generate any pitch by adjusting the amplitude and number of overtones and undertones accordingly. The following example computes a single tone conforming to given spectral centroid and spectral spread (weighted variance) values. For an odd number of natural harmonics with an identical gain for non-central oscillators,

$$\begin{aligned} &\text{for } N \in \{1, 3, 5, \dots\}, \\ &n = 1, 2, 3, \dots, N, \\ &a_n = \begin{cases} 1 & \text{when } n = \frac{N+1}{2} \\ g & \text{otherwise} \end{cases} \\ &g = \frac{\sigma^2}{\sum (v_n - \mu)^2 + \sigma^2(1-N)}, \end{aligned} \quad (5)$$

where  $a$  is the gain coefficients with a symmetric form  $\{\dots, g, g, 1, g, g, \dots\}$  and  $\sigma$  is the square root of the spectral spread (i.e., standard deviation, Online Example 12). Similarly, we can construct an arbitrary harmony with sinusoidal oscillators centered around a given spectral centroid, such that

$$v_n = \sum_{n=0}^{N-1} \frac{\mu C_n}{\bar{C}}, \quad (6)$$

where  $C$  is a vector of normalized frequency coefficients in Hz for creating a chord<sup>2</sup> and  $\bar{C}$  is the average frequency of them (Online Example 13). Combining both additive synthesis and harmony, it is also feasible to generate any chord on any root note from given spectral parameters (Online Example 14). The flexibility in generating the pitch or the chord quality can be utilized to encode additional data dimensions for human perception. These harmonic techniques are relatively robust for retrieving the spectral parameters, provided that the data mapped to the spectral centroid is scaled properly so that the harmonic or chord-voice frequencies exist within the spread range.

#### 4.1.3. Further Applications

Lastly, in addition to encoding data to multiple spectral parameters for generating a single spectral distribution, the potential of SPE is the ability to mix multiple timbral techniques (e.g., harmony model + single-additive-voice + noise percussion) as long as they conform to the overall distribution parameters (Online Example 15), or even to mix multiple distributions and estimate their parameters with a mixture-model parameter estimation [22]. Mixing multiple instruments to a single distribution enables additional

<sup>2</sup> This may be obtained by converting a list of degrees in MIDI note number starting at 0 (e.g., [0, 4, 7, 11] for the major seventh) to frequency.

perceptual dimensions that may be tracked by the human listener (e.g., the salience of a particular instrument), while preserving the most critical data channels in the spectral parameters for computational recovery.

## 5. DISCUSSION AND FUTURE WORK

To summarize, SPE encapsulates the data points, either from raw input or analyzed structures, to the abstract statistical shape or parameters (e.g., mean and variance) of a magnitude-spectrum frame. This facilitates a uniquely constrained yet flexible composition expanded from the target magnitude spectrum, and even allows additional mapping of data to such as the choice of chord or onset shape for perceptual decoding. As we include multiple data dimensions in the analysis, mapping, and encoding paths, the entire signal flow may grow into a quite complex system as Figure 4.

We did not, however, examine musical expressions that extend over several seconds (e.g., rhythms, melodic patterns) in this discussion. For future work, we plan on examining spectral encoding techniques over time utilizing symbolic parameterizations. Relevant work includes Smalley's spectromorphology [23], an analytical framework for electro-acoustic music in which the author lists qualitative distinctions in each morphing (moving) steps of spectral contents. The chosen parameters (e.g., "upbeat" + "transition" + "closure") combine and form a complex musical gesture over time. In addition, spectral modeling synthesis [18] also provides insights in creating time-varying timbral structure with the deterministic and random components.

The time-varying encoding poses a practical issue with the time resolution of the data stream. SPE analyzes the STFT frames of the output audio with a reasonable frequency resolution (e.g., 1024 samples at the sampling rate of 44100 for harmonic or granular approach), which limits the data rate to at least one datum per 20 milliseconds. This is quite slow compared to, for example, audification or even possibly PMS. The data rate is forced to decrease even more when using encoding techniques such as granular synthesis or mixed-timbre composition because of the susceptibility to voice phasing. Also, adding time-domain audio effects such as delay and reverberation also smears out the phase relationship, causing more errors in machine listening.

## 6. CONCLUSION

We presented spectral parameter encoding, a dual-layer framework for musically expressive yet functional design of sonification. It employs a simple structural analysis to facilitate a semi-automated organization of mapping, and data encoding to spectral features as well as computational feature extraction to ensure the minimized loss of information as a whole in the process of transformation and mapping. Although the use of spectral distribution imposes certain acoustic constraints, it allows a variety of musically-organized sonification from timbral to harmonic expressions with the possibility of a multi-timbral structure.

## 7. REFERENCES

[1] P. Vickers and B. Hogg, "Sonification Abstraite/Sonification Concrete: An 'Aesthetic Persepective Space' for Classifying Auditory Displays in the Ars Musica Domain," in *Proceedings of the 12th*

*International Conference on Auditory Display (ICAD 2006)*, 2006.

- [2] C. Roads, *Microsound*. Cambridge: MIT Press, 2001.
- [3] N. Degara, F. Nagel, and T. Hermann, "Sonex: An Evaluation Exchange Framework For Reproducible Sonification," Jul. 2013.
- [4] K. Vogt, "A Quantitative Evaluation Approach to Sonifications," Jun. 2011.
- [5] A. Klapuri and M. Davy, *Signal Processing Methods for Music Transcription*. New York: Springer, 2006.
- [6] A. Lerch, *Audio Content Analysis: An Introduction*. Hoboken, N.J.: Wiley, 2012.
- [7] B. Logan, "Mel Frequency Cepstral Coefficients for Music Modeling."
- [8] F. Grond and J. Berger, "Parameter Mapping Sonification," *Sonification Handb.*, pp. 363–397, 2011.
- [9] T. Hermann and H. Ritter, "Listen to your Data: Model-Based Sonification for Data Analysis," in *189–194, Int. Inst. for Advanced Studies in System research and cybernetics*, 1999, pp. 189–194.
- [10] T. Hermann, "Model-Based Sonification," *Sonification Handb.*, pp. 399–427, 2011.
- [11] R. L. Alexander, J. A. Gilbert, E. Landi, M. Simoni, T. H. Zurbuchen, and D. A. Roberts, "Audification as a Diagnostic Tool for Exploratory Heliospheric Data Analysis," Jun. 2011.
- [12] A. Tanaka, "The Sound of Photographic Image," *AI Soc.*, vol. 27, no. 2, pp. 315–318, May 2012.
- [13] T. Hermann, P. Meinicke, and H. Ritter, "Principal Curve Sonification," 2000.
- [14] T. Hermann, P. Meinicke, H. Bekel, H. Ritter, H. M. Müller, and S. Weiss, "Sonification for EEG Data Analysis," in *Proceedings of the 2002 International Conference on Auditory Display*, 2002.
- [15] F. Djebbar, B. Ayad, K. A. Meraim, and H. Hamam, "Comparative Study of Digital Audio Steganography Techniques," *EURASIP J. Audio Speech Music Process.*, vol. 2012, no. 1, p. 25, 2012.
- [16] T. Tsuchiya, J. Freeman, and L. W. Lerner, "Data-to-Music API: Real-Time Data-Agnostic Sonification with Musical Structure Models," *Proc 21st Int Conf Audit. Disp.*, 2015.
- [17] M. Malt and E. Jourdan, "Zsa.Descriptors: a library for real-time descriptors analysis," *ResearchGate*.
- [18] X. Serra and J. Smith, "Spectral Modeling Synthesis: A Sound Analysis/Synthesis System Based on a Deterministic Plus Stochastic Decomposition," *Comput. Music J.*, vol. 14, no. 4, pp. 12–24, 1990.
- [19] T. Tsuchiya, J. Freeman, and L. W. Lerner, "Data-Driven Live Coding with DataToMusic API," in *Proceedings of the 2nd Web Audio Conference (WAC-2016)*, Atlanta, 2016.
- [20] R. Bracewell, *The Fourier Transform and Its Applications*, vol. 5. 1965.
- [21] S. Olver and A. Townsend, "Fast Inverse Transform Sampling in One and Two Dimensions," *ArXiv13071223 Math Stat*, Jul. 2013.
- [22] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker Verification Using Adapted Gaussian Mixture Models," *Digit. Signal Process.*, vol. 10, no. 1, pp. 19–41, Jan. 2000.
- [23] D. Smalley, "Spectromorphology: Explaining Sound-Shapes," *Organised Sound*, vol. 2, no. 2, pp. 107–126, Aug. 1997.